

DIALOG(R) File 347:JAPIO
(c) 2004 JPO & JAPIO. All rts. reserv.

02969070 **Image available**
EXTRACTING PROCESSING SYSTEM FOR CHARACTERISTIC VOCABULARY IN
JAPANESE
OBJECT SENTENCE

PUB. NO.: 01-266670 [JP 1266670 A]
PUBLISHED: October 24, 1989 (19891024)
INVENTOR(s): OKU MASAHIRO
 HIGASHIDA MASANOBU
APPLICANT(s): NIPPON TELEGR & TELEPH CORP <NTT> [000422] (A Japanese
 Company or Corporation), JP (Japan)
APPL. NO.: 63-095096 [JP 8895096]
FILED: April 18, 1988 (19880418)
INTL CLASS: [4] G06F-015/38
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications);
30.2
 (MISCELLANEOUS GOODS -- Sports & Recreation)
JOURNAL: Section: P, Section No. 992, Vol. 14, No. 27, Pg. 20,
January
 19, 1990 (19900119)

ABSTRACT

PURPOSE: To automatically extract characteristic vocabulary in an object sentence by classifying a Japanese document into character type code strings, extracting the candidates of object characteristic vocabulary from the code strings, further extracting the candidates of the object characteristic vocabulary having high accuracy out of all the above-mentioned candidates based on language information, and further outputting vocabulary which does not exist in a Japanese dictionary for analysis.

CONSTITUTION: For the inputted Japanese document, a code string expanding part 1 generates plural types of the character type code strings for every character in the Japanese document. A characteristic vocabulary candidate extracting part 2 extracts all the character strings corresponding to the code strings in an extracting character type string prescribing table 7 as the candidates of the characteristic vocabulary in the Japanese object sentence and classifies the candidates according to conditions in a classifying table 8. Next, a characteristic vocabulary language

BEST AVAILABLE COPY

processing
part 3 retrieves a language information table 9, processes
respective
above-mentioned candidates, and extracts some candidates out of the
above-
mentioned candidates having higher accuracy. A characteristic
vocabulary
language selecting part 4 retrieves a dictionary 10 for the analysis
with
the shapes of the characters of the candidates from the processing
part 3
as keys and removes the candidates from all the candidates when
the
candidates are already registered in the dictionary 10. The candidates
are
regarded as the characteristic vocabulary in the Japanese object
sentence,
sent to a registering part, and written and registered into a file 6
when
the candidates are not registered yet.
?

特公平8-10452

(24) (44) 公告日 平成8年(1996)1月31日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/27		8420-5L	G 0 6 F 15/ 38	E

請求項の数1(全 11 頁)

(21) 出願番号	特願昭63-95096	(71) 出願人	999999999 日本電信電話株式会社 東京都新宿区西新宿3丁目19番2号
(22) 出願日	昭和63年(1988)4月18日	(72) 発明者	奥 雅博 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
(65) 公開番号	特開平1-266870	(72) 発明者	東田 正信 東京都千代田区内幸町1丁目1番6号 日 本電信電話株式会社内
(43) 公開日	平成1年(1989)10月24日	(74) 代理人	弁理士 森田 寛
		審査官	鶴谷 裕二
		(56) 参考文献	特開 平1-79863 (J P, A)

(54) 【発明の名称】 日本語対象文固有用語抽出処理装置

【特許請求の範囲】

【請求項1】 日本語文書中の文字列を複数種類の字種に分類したコード列に展開する第1の手段、その文書でのみ使用されている製品名や会社名や人名を含む固有名詞と単語の組み合わせであっても新語やその文書でのみ使用されていると考えられる語とを総称する日本語対象文固有用語について、前記コード列における字種の変化点および並び方のみから当該日本語対象文固有用語の候補を抽出する第2の手段、日本語の性質から、前記日本語対象文固有用語とはならない文字列の条件と当該日本語対象文固有用語になりやすい文字列の条件とを記憶した言語情報テーブルを用いて、当該日本語対象文固有用語になりやすい文字列の条件を満たす候補のみを残す第3の手段、および第3の手段で抽出された候補の中から、日本語単

語の字面や品詞を含む情報が登録されている日本語辞書に収録されていない語のみを前記日本語対象文固有用語として出力する第4の手段を有することを特徴とする日本語対象文固有用語抽出処理装置。

【発明の詳細な説明】

(1) 発明の属する技術分野

本発明は、日本語文書中から、該文書でのみ使用されている製品名、会社名、人名などの固有名詞や単語の組み合わせであっても新語や該文書においてのみ使用されていると考えられる語（対象文固有用語）を自動的に抽出する日本語対象文固有用語抽出処理装置に関するものである。

(2) 従来技術

従来から文作成に当って、日本語文書中から、該文書

でのみ使用されている製品名、会社名、人名などの固有名詞や、単語の組み合わせであっても新語や、該文書においてのみ使用されていると考えられる語（対象文固有用語）を抽出することが行われている。このための従来の装置では、入力日本語をあらかじめ用意した解析辞書を用いて分かち書きを行い、その結果、うまく分かち書きできなかった部分を対象文固有用語として抽出する。また、字種の変化点に着目する装置では、ある字種から別の字種へ変化する点までを対象文固有用語の候補として抽出し、該候補の中から、あらかじめ用意した解析用の日本語辞書に登録されていない候補を、対象文固有用語として抽出するようにしている。

しかし、前者の装置では、

① 日本語の分かち書き処理は、解析辞書中に単語が登録されていることを前提としているため、対象文固有用語についても解析を試み、部分的に解析に成功すると、その部分を対象文固有用語の対象外となる（例：「〇×会社」は全体で社名を表わす対象文固有用語であるにもかかわらず、「会社」は一般的な用語であって、解析に成功するため、「〇×」のみが対象文固有用語であると認定されてしまう）など、正確に対象文固有用語を抽出することができない。

② 対象文固有用語が一般的な語によって構成されている場合、該用語は、対象文固有用語として認定されない。例を以下に示す。「日本電信電話株式会社」は固有名詞であり、対象文固有用語であるが、一般的な語のみによって構成されているために、「日本／電信／電話／株式／会社」のように分かち書きに成功する。このため、対象文固有用語として抽出されるべき「日本電信電話株式会社」が抽出されない。

また、後者の装置では、

- ① 複数の字種にわたる対象文固有用語が抽出しづらい。
- ② 「～向け」などの送りがない付きの接辞を含む対象文固有用語が抽出できない。
- ③ 途中にひらがなを含む対象文固有用語（例：係り受け解析）を抽出できない。

（３）発明の目的

本発明の目的は、前記の問題点を解決した、日本語文書中から対象文固有用語を正確に自動抽出する日本語対象文固有用語抽出処理装置を提供することにある。

（４）発明の構成

（４－１）発明の特徴と従来の技術との差異

本発明は、日本語文書中の文字列を複数種類の字種に分類したコード列に展開する第１の手段、

その文書でのみ使用されている製品名や会社名や人名を含む固有名詞と単語の組み合わせであっても新語やその文書でのみ使用されていると考えられる語とを総称する日本語対象文固有用語について、前記コード列における字種の変化点および並び方のみから当該日本語対象文

固有用語の候補を抽出する第２の手段、

日本語の性質から、前記日本語対象文固有用語とはならない文字列の条件と当該日本語対象文固有用語になりやすい文字列の条件とを記憶した言語情報テーブルを用いて、当該日本語対象文固有用語になりやすい文字列の条件を満たす候補のみを残す第３の手段、

および第３の手段で抽出された候補の中から、日本語単語の字面や品詞を含む情報が登録されている日本語辞書に収録されていない語のみを前記日本語対象文固有用語として出力する第４の手段を有することを特徴とする。

従来の技術とは、

- ① 第２の手段において、字種の変化点だけでなく、並び方にも着目して候補を抽出するので、ひらがなを含む日本語対象文固有用語も候補として抽出できる。
- ② 第２の手段において、字種の変化点および並び方に着目して候補を抽出するので、一般的な語のみから成る語や一部に一般的な語を含む語も日本語対象文固有用語の候補として抽出できる。
- ③ 第３の手段において、「漢字５文字以上の列は日本語対象文固有用語となりやすい」などのヒューリスティックルールや、「該候補中に接辞が含まれている場合には、接辞に対する処理を施す」などの処理情報から成る言語情報を用いて、日本語対象文固有用語の候補を絞るので、正確な候補抽出が行える。
- ④ 第４の手段においては、解析辞書中の単語と完成一致しない語はすべて日本語対象文固有用語として出力するため、一般的な語のみからなる日本語対象文固有用語や、一部に一般的な語を含む日本語対象文固有用語も、日本語対象文固有用語として抽出できる。
- の各点が異なる。

（４－２）実施例

第１図は本発明をハードウェアによって構成した際の基本構成図である。図中、１は、入力された日本語文書を例えば１０種類の字種（漢字コード、漢数字コード、一般ひらがなコード、連用形活用語尾ひらがなコード、カタカナコード、アラビア数字コード、アルファベットコード、句読点コード、区切りコード、その他のコード）に分類したコード列に展開するコード列展開部。２は、コード列展開部１において得られたコード列の中から、字種の変化点および並び方に着目して、日本語対象文固有用語の候補を抽出する日本語対象文固有用語候補抽出部。３は、日本語対象文固有用語候補抽出部２において得られた候補の中の各候補に対して言語情報テーブル９を検索し、該情報に従って該候補を処理した後、より精度の高い候補のみから成る候補群を抽出する日本語対象文固有用語言語処理部。４は、日本語対象文固有用語言語処理部３において得られた候補群の中の各候補の字面をキーとして日本語解析辞書１０を検索し、日本語解析辞書１０に登録されていない語のみを日本語対象文固有用語

として選択する日本語対象文固有有用語選択部。5は、日本語対象文固有有用語選択部4で選択された日本語対象文固有有用語を日本語対象文固有有用語ファイル6に登録する日本語対象文固有有用語登録部。6は、最終的に抽出された日本語対象文固有有用語を登録しておく日本語対象文固有有用語ファイル。7は、候補として抽出すべき字種の並びを規定した抽出字種列規定テーブル。8は、日本語対象文固有有用語候補抽出部2において抽出する字種列をどのように分類するかを規定した分類テーブル。9は、日本語対象文固有有用語言語処理部3において、より精度の高い候補を抽出する際に用いる言語情報や処理方法を記述した言語情報テーブル。10は、一般的な日本語単語の字面や品詞などを登録した日本語解析辞書。11は、演算装置およびメモリから成る日本語対象文固有有用語抽出装置である。

第2図は第2図(A)と(B)とで一体となって1つの図を構成するものであり、日本語対象文固有有用語抽出装置11の動作の概略フローである。

次に、第2図の概略フローに従って、動作の説明を行う。

日本語対象文固有有用語抽出装置11の入力である日本語文書に対して、コード列展開部1では、該日本語文書の1文字1文字を10種類の字種(漢字コード、漢数字コード、一般ひらがなコード、連用形活用語尾ひらがなコード、カタカナコード、アラビア数字コード、アルファベットコード、句読点コード、区切りコード、その他のコード)のコードに変換し、該日本語文書に対するコード列を生成する(第2図ステップ1)。

このとき、各コードには、該コードが該日本語文書のどの文字から生成されたのかを示す情報が付与される。コード列展開部1によって生成されたコード列は、日本語対象文固有有用語候補抽出部2に送られる。

日本語対象文固有有用語候補抽出部2では、抽出字種列規定テーブル7に記述されているコード列に対応する文字列を、日本語文書中から、日本語対象文固有有用語の候補としてすべて抽出し、分類テーブル8に記述されている条件によって該候補を分類する(第2図ステップ2)。

分類された日本語対象文固有有用語の後方は、日本語対象文固有有用語言語処理部3に送られる。第3図に、分類テーブル8の内容の例を示す。日本語対象文固有有用語言語処理部3では、まず、言語情報テーブル9を検索して情報を得る(第2図ステップ31)。次に、得られた情報に従って、日本語対象文固有有用語の各候補について処理する(第2図ステップ32)ことによって、より精度の高い日本語対象文固有有用語の候補を抽出し(第2図ステップ33)、日本語対象文固有有用語選択部4に該候補を送る。

日本語対象文固有有用語選択部4では、日本語対象文固有有用語言語処理部3により送られてきた日本語対象文

固有語の候補の字面をキーとして、日本語解析辞書10を検索する。検索の結果、該候補が日本語解析辞書10に登録されているときは、該候補は日本語対象文固有有用語ではないとして、候補から落とす(第2図ステップ4)。逆に、該候補が日本語解析辞書10に登録されていないときは、該候補は日本語対象文固有有用語であるとして、該候補を日本語対象文固有有用語登録部5に送る(第2図ステップ4)。

日本語対象文固有有用語登録部5では、日本語対象文固有有用語選択部4より送られてきた日本語対象文固有有用語を日本語対象文固有有用語ファイル6に書き込み、登録する(第2図ステップ5)。

次に例を用いて動作の概略を説明する。第4図に示す例文と日本語対象文固有有用語抽出装置11の入力となる日本語文書として説明する。

まず、コード列展開部1では、第4図に示す例文の1文字1文字を対応するコードに変換し、コード列を生成する(第5図)。第5図のコード列では、漢字コード(漢数字を除く全ての漢字)を(), 漢数字コード(〇, 一, 二, 三, 四, 五, 六, 七, 八, 九)を(), 一般ひらがなコード(連用形活用語尾ひらがなコード以外の全てのひらがな)を(), 連用形活用語尾ひらがなコード(い, き, し, ち, み, り, え, け, せ, て, ね, め, れ, ぎ, じ, び, げ, ぜ, べ, っ)を(), カタカナコード(カタカナすべて)を(), アラビア数字コード(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)を(), アルファベットコード(A~Z, a~zの大文字, 小文字)を(), 句読点コード(読点, 句点, カンマ, ビリオド, クエスションマーク, イクスクラメーションマーク)を(), 区切りコード(カギカッコ, 丸カッコなどのカッコ類, コーテーションマーク, ダブルコーテーションマーク)を(), その他のコード(前記のどのコードには入らない文字や記号)を()と略記している。このとき、各コードには、該コードがどの文字から生成されたかを示す情報(例えば、第5図の一番最初の()は第4図の文字“位”から、次の()は文字“置”からそれぞれ生成されたことを示す情報)を付与する。この付与の方法には、元の文字列に番号を付け、その番号をコード列に付与する方法や、元の文字とコードをペアで持つ方法などがあるが、ここでは、この方法については問わない。

コード列展開部1によって生成されたコード列(第5図)は、日本語対象文固有有用語候補抽出部2に送られる。

次に、日本語対象文固有有用語候補抽出部2では、第5図に示したコード列から1つ1つのコードを取り出し、抽出字種列規定テーブル7中のルールと照合し、入力日本語文書からすべての候補を抽出する。第6図に抽出字種列規定テーブル7の内容例をif~thenルールの形で表わしたものを示す。例文において、まず、第1文字目の“位”に対するコード()が処理される。このコードは、

抽出字種列規定テーブル7における現在のコード=()のルールと一致するので、このコード()に対応する文字“位”が候補文字列バッファに加えられる(第7図(1)参照)。

次に、第2文字目の“置”に対するコード(), 第3文字目の“繰”に対するコード()が同様に処理され(第7図(2), (3)), ここまでの処理で候補文字列バッファには第7図(3)に示す“位置繰”が入っている。次に第4文字目の“り”に対するコード()が処理される。このコードは抽出字種列規定テーブル7における現在のコード()のルールと一致する。このルールは次の文字に対するコードによってさらに処理が分かれている。次の文字は第5文字目の“返”であり、対応するコードは()である。従って抽出字種列規定テーブル7中の該ルールにより、現在のコード()に対応する文字“り”が候補文字列バッファに加えられる(第7図(4))。次に第5文字目が処理され、候補文字列バッファの内容は第7図(5)のようになる。同様に、第8文字目の“度”までが候補文字列バッファ内に加えられる(第7図(8))。次に第9文字目の“は”に対するコード()が処理される。このコードは抽出字種列規定テーブル7における現在のコード=()のルールに一致する。このルールにより、今までの処理で候補文字列バッファに入っている文字列「位置繰り返し精度」が候補として抽出され、候補文字列バッファがクリアされる。以上のような動作を繰り返し、入力日本語文書の最後まで処理を行い、すべての候補が抽出される(第8図、第9図参照)。

さらに、日本語対象文固有用語候補抽出部2では、抽出した候補を分類テーブル8に従って分類する。この結果を第10図に示す。分類結果は日本語対象文固有用語言語処理部3に送られる。

日本語対象文固有用語言語処理部3では、言語情報テーブル9を検索し、得られた情報に従って候補を絞り込む。言語情報テーブル9の内容の例を第11図に示す。日本語対象文固有用語言語処理部3では、第10図に示す候補の分類をキーとして、言語情報テーブル9を検索する。第10図の分類1に属する語「上」は、第11図分類1の項の「無条件に候補から落とす」に一致するため、候補から落とされる。また、分類23に属する「〇・〇五ミリ」、「毎秒二・ニメートル」の2語は、第11図分類23の項の「漢数字とそれに続く助数詞を持つものは候補から落とす」に一致するため、候補から落とされる。分類2の「性能」、分類3の「在来機」、分類6の「最大合成速度」、分類22の「位置繰り返し精度」の4語は候補として残り、日本語対象文固有用語選択部4に送られる。

日本語対象文固有用語選択部4では、日本語解析辞書10を日本語対象文固有用語の候補の字面で検索し、日本語解析辞書10に登録されていない語のみを日本語対象文

固有用語として選択する。●に対しては、上記4語の候補の各々について、その字面をキーとして日本語解析辞書10を検索する。検索の結果、日本語解析辞書には、「性能」のみが一般的な語として登録されているため、「性能」が候補から落とされる。よって上記の4語から「性能」を除いた3語が日本語対象文固有用語として日本語対象文固有用語登録部5に送られる。

日本語対象文固有用語登録部5では、送られてきた3個の日本語対象文固有用語を日本語対象文固有用語ファイル6に書き込み、登録する。日本語対象文固有用語ファイル6に書き込まれた日本語対象文固有用語を第12図に示す。

このような構造および作用となっていることから、従来の方法に比べて、

① 日本語対象文固有用語の候補を抽出する際に、字種の変化点だけでなく、並び方にも着目しているため、ひらがなを含む日本語対象文固有用語も候補として抽出できる。

② 字種の変化点および並び方に着目して候補を抽出するので、一般的な語のみから成る語や一部に一般的な語を含む語も日本語対象文固有用語の候補として抽出できる。

③ 字種の変化点および並び方に着目して抽出した候補に対して、言語情報テーブル中の情報(「漢字5文字以上の列は日本語対象文固有用語となりやすい」といったヒューリスティックルールや、「接辞を含む場合には、接辞に対する処理を施す」などの処理情報)を用いて、候補を絞り込むので、正確な日本語対象文固有用語の候補抽出が行える。

④ 日本語解析辞書中の単語と完全一致しない語はすべて日本語対象文固有用語として抽出するため、一般的な語のみから成る日本語対象文固有用語や、一部に一般的な語を含む日本語対象文固有用語も、日本語対象文固有用語として抽出できる。

の各点で改善があった。

(5) 発明の効果

以上、説明したように、本発明によれば、該文書でのみ使用されている製品名、会社名、人名などの固有名詞や単語の組み合わせであっても新語や該文書においてのみ使用されていると考えられる語(以上の語をまとめて日本語対象文固有用語と呼ぶ)の候補を該文書中から、字種の変化点および並び方の情報によって抽出し、該宝庫の持つ言語情報を用いて候補を絞った後に、解析辞書を検索することによって一般的な語を取り除いたものを日本語対象文固有用語として出力するのであるから、日本語文書中に存在する日本語対象文固有用語を、正確に抽出できるという利点がある。

【図面の簡単な説明】

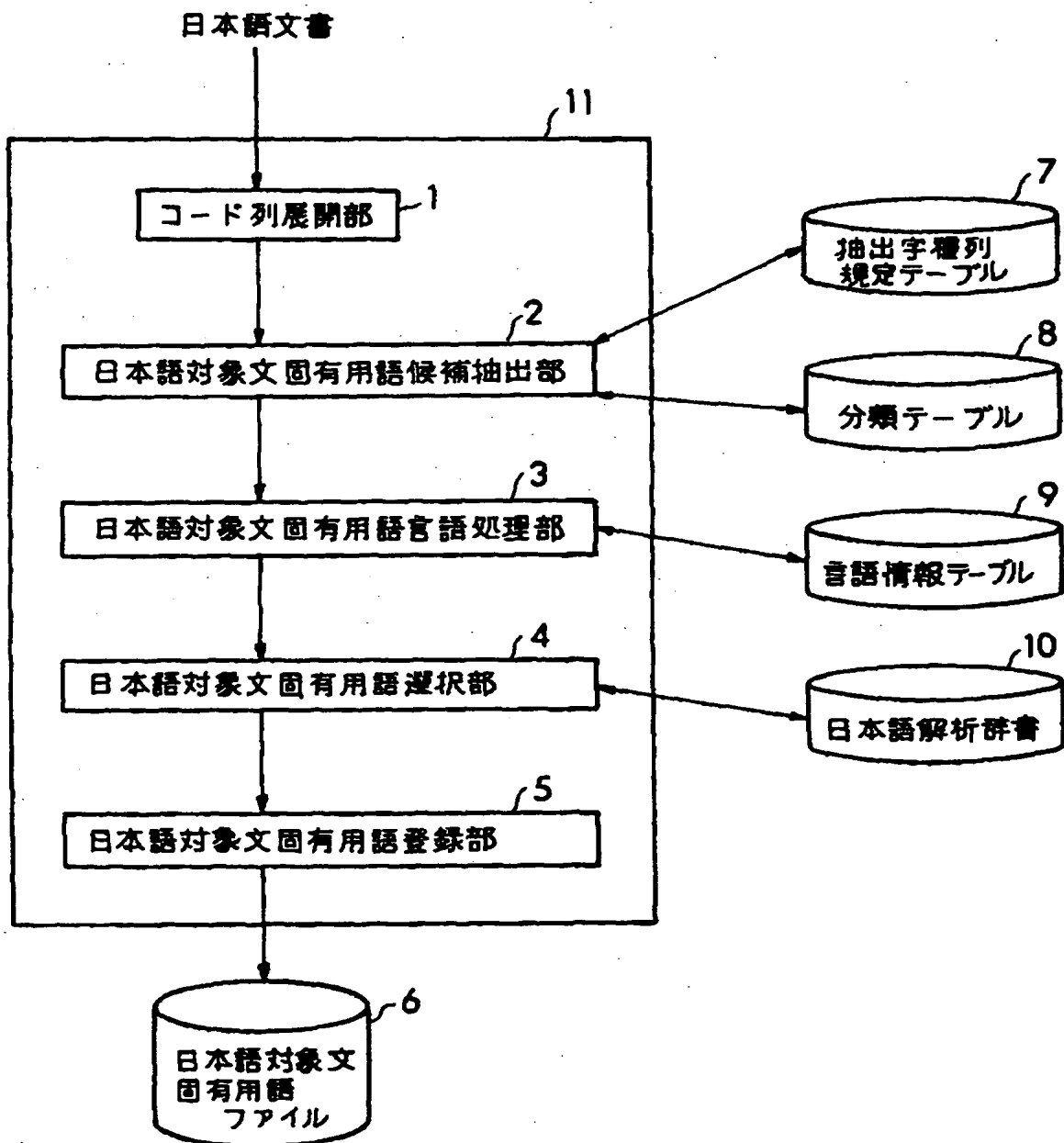
第1図は本発明の基本構成図、第2図は本発明の動作の概略フロー、第3図は分類テーブルの内容の例、第4図

は動作の説明に用いた例文、第5図は例文に対するコード列、第6図は抽出字種列規定テーブルの内容例、第7図は候補文字列バッファ内の動作例、第8図は第5図のコード列から抽出される日本語対象文固有用語の候補のコード列、第9図は第8図のコード列に対応する元の文字列、第10図は分類テーブルによる日本語対象文固有用語の候補の分類、第11図は言語情報テーブルの内容の例、第12図は最終的に抽出された日本語対象文固有用語

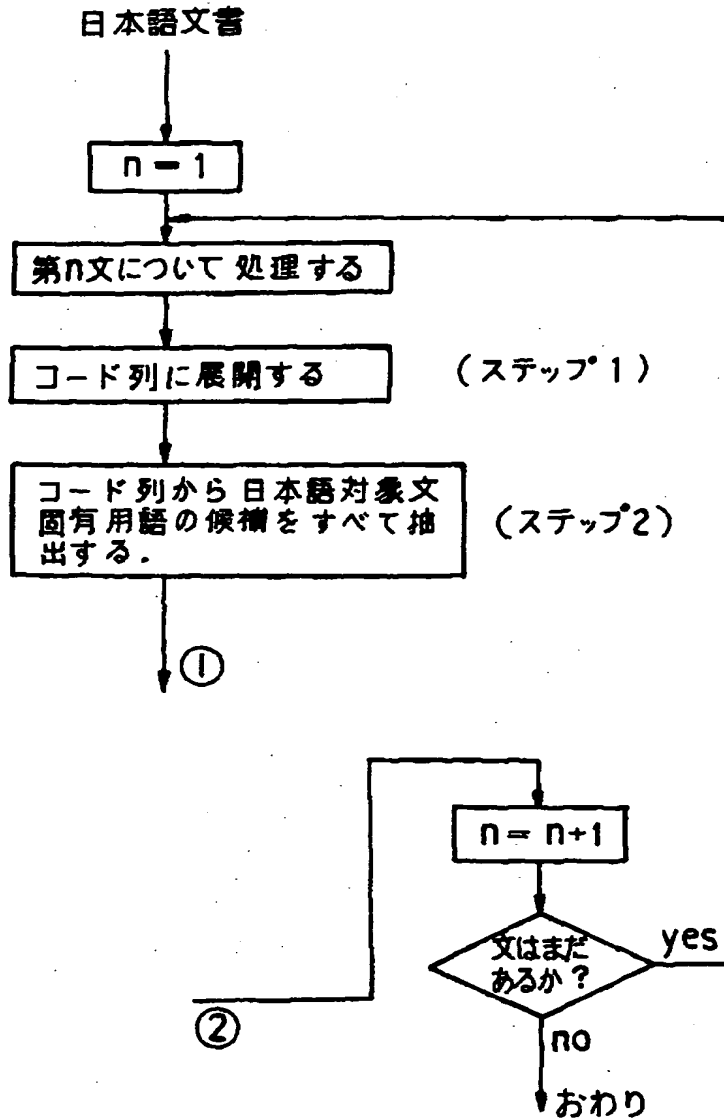
を示す。

1…コード列展開部、2…日本語対象文固有用語候補抽出部、3…日本語対象文固有用語言語処理部、4…日本語対象文固有用語選択部、5…日本語対象文固有用語登録部、6…日本語対象文固有用語ファイル、7…抽出字種列規定テーブル、8…分類テーブル、9…言語情報テーブル、10…日本語解析辞書、11…日本語対象文固有用語抽出装置。

【第1図】



基本構成図



動作の概略フロー

【第4図】

位置繰り返し精度は〇・〇五ミリ、最大合成速度も
毎秒二・ニメートルと在来機よりも性能が上がった。

例 文

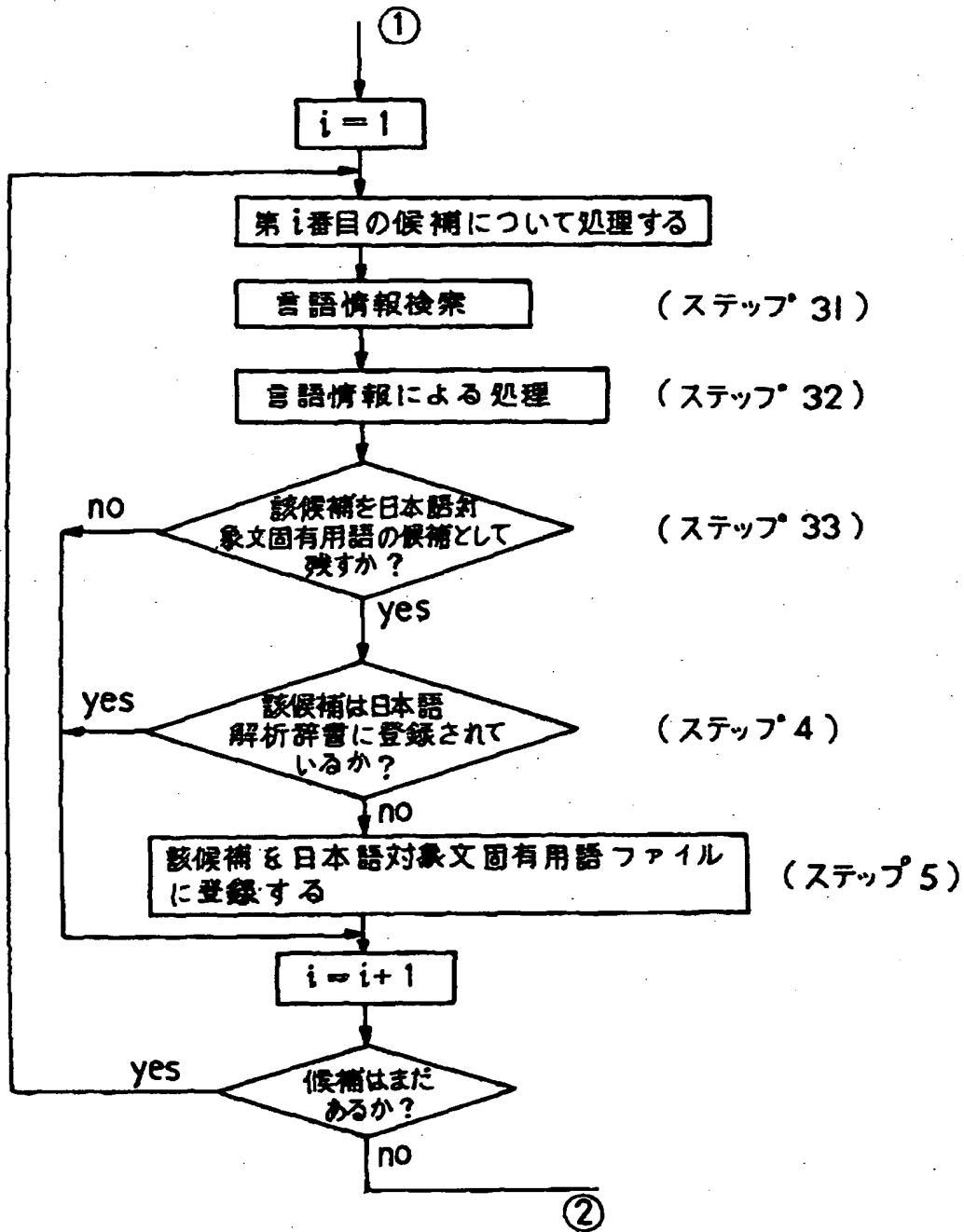
- (1) 候補文字列バッファ 位
- (2) 候補文字列バッファ 位置
- (3) 候補文字列バッファ 位置繰
- (4) 候補文字列バッファ 位置繰り
- (5) 候補文字列バッファ 位置繰り返
- (6) 候補文字列バッファ 位置繰り返し
- (7) 候補文字列バッファ 位置繰り返し精
- (8) 候補文字列バッファ 位置繰り返し精度

候補文字列バッファ内の動作例

【第12図】

分類番号	日本語対象文固有用語
分類 3	在 来 機
分類 6	最大合成速度
分類 22	位置繰り返し精度

例文に対して抽出された日本語対象文固有用語



動作の概略フロー

【第3図 (A)】

分類番号	対象となる文字列の条件
分類1	漢字1文字からなる文字列
分類2	漢字2文字からなる文字列
分類3	漢字3文字からなる文字列
分類4	漢字4文字からなる文字列
分類5	漢字5文字からなる文字列
分類6	漢字6文字以上からなる文字列
分類7	漢字のみからなる文字列にカタカナのみからなる文字列が接続している文字列
分類8	漢字のみからなる文字列にアルファベットのみからなる文字列が接続している文字列
分類9	漢字のみからなる文字列にアラビア数字のみからなる文字列が接続している文字列
分類10	カタカナのみからなる文字列に漢字のみからなる文字列が接続している文字列
分類11	カタカナのみからなる文字列にアルファベットのみからなる文字列が接続している文字列
分類12	カタカナのみからなる文字列にアラビア数字のみからなる文字列が接続している文字列

分類テーブルの内容

【第3図 (B)】

分類13	アルファベットのみからなる文字列に漢字のみからなる文字列が接続している文字列
分類14	アルファベットのみからなる文字列にカタカナのみからなる文字列が接続している文字列
分類15	アルファベットのみからなる文字列にアラビア数字のみからなる文字列が接続している文字列
分類16	アラビア数字のみからなる文字列に漢字のみからなる文字列が接続している文字列
分類17	アラビア数字のみからなる文字列にカタカナのみからなる文字列が接続している文字列
分類18	アラビア数字のみからなる文字列にアルファベットのみからなる文字列が接続している文字列
分類19	カタカナのみからなる文字列
分類20	アルファベットのみからなる文字列
分類21	アラビア数字のみからなる文字列
分類22	ひらがなを含む文字列
分類23	上記以外の文字列

注1) 上記の「漢字」には、「漢数字」も含まれる。
 注2) 分類7～分類18までの文字列の後ろには別の文字列が接続していても良い。

分類テーブルの内容

【第5図】

(A)(A)(A)(D)(A)(D)(A)(A)(C)(B)(J)(B)(B)(E)(E)(H)(A)(A)(A)(A)(A)(C)
 (A)(A)(B)(J)(B)(E)(E)(E)(E)(C)(A)(A)(A)(C)(D)(C)(A)(A)(C)(A)(C)(D)(C)(H)

漢字コード …… (A), 漢数字コード …… (B),
 一般ひらがなコード …… (C), 連用形活用語尾 ひらがなコード …… (D),
 カタカナコード …… (E), アラビア数字コード …… (F),
 アルファベットコード …… (G), 句読点コード …… (H),
 区切りコード …… (I), その他のコード …… (J)

例文に対するコード列

【第6図】

```

if 現在のコード = (A) then (A) に対応する文字列を候補文字列バッファに加える
if 現在のコード = (B) then (B) に対応する文字列を候補文字列バッファに加える
if 現在のコード = (C) then 候補文字列バッファ内の文字列を候補とし、
                           候補文字列バッファをクリアする
if 現在のコード = (D) then
    if 次のコード = (A) or (B) or (E) or (F) or (G) or (I) then
        (D) に対応する文字列を候補文字列バッファに加える
    else
        候補文字列バッファ内の文字列を候補とし、候補文字列をクリアする
if 現在のコード = (E) then (E) に対応する文字列を候補文字列バッファに加える
if 現在のコード = (F) then (F) に対応する文字列を候補文字列バッファに加える
if 現在のコード = (G) then (G) に対応する文字列を候補文字列バッファに加える
if 現在のコード = (H) then 候補文字列バッファ内の文字列を候補とし、
                           候補文字列バッファをクリアする
if 現在のコード = (I) then 候補文字列バッファ内の文字列を候補とし、
                           候補文字列バッファをクリアする
if 現在のコード = (J) then (J) に対応する文字列を候補文字列バッファに加える
    
```

抽出字種列規定テーブルの内容例

【第8図】

A A A D A D A A C
B J B B E E H
A A A A A A C
A A B J B E E E E
C A A A C D C
A A C A C D C H

..... 抽出字種別規定テーブルの記述に
 従って抽出された日本語対象文
 図有用語の候補

コード列から抽出される日本語対象文図有用語の候補

【第9図】

位置繰り返し精度	〇・〇五ミリ	最大合成速度
毎秒二・ニメートル	在来機	性能
上		

第8図の枠で囲ったコード列に対する元の文字列

【第11図】

文字列の属する分類	処 理
分類1	・無条件に候補から落とす。
分類2	・漢数字のみから成るものは候補から落とす。 ・副詞的に使われる名詞（今日など）は候補から落とす。 ・漢数字とそれに続く助数詞を持つものは候補から落とす。
分類3	・漢数字とそれに続く助数詞を持つものは候補から落とす。 ・漢数字のみから成るものは候補から落とす。 ・副詞的に使われる名詞を含むものは候補から落とす。
分類4	・漢数字とそれに続く助数詞を持つものは候補から落とす。 ・漢数字のみから成るものは候補から落とす。 ・副詞的に使われる名詞を含むものは候補から落とす。
分類5	・漢数字とそれに続く助数詞を持つものは候補から落とす。 ・漢数字のみから成るものは候補から落とす。
分類6	・漢数字とそれに続く助数詞を持つものは候補から落とす。 ・漢数字のみから成るものは候補から落とす。
⋮	⋮
分類20	・無条件に候補とする。
分類21	・無条件に候補から落とす。
分類22	・無条件に候補とする。
分類23	・漢数字とそれに続く助数詞を持つものは候補から落とす。

言語情報テーブルの内容の例

【第10図】

分類番号	抽出される日本語対象文固有用語の候補
分類1	上
分類2	性能
分類3	在来機
分類4	
分類5	
分類6	最大合成速度
分類7	
分類8	
分類9	
分類10	
分類11	
分類12	
分類13	
分類14	
分類15	
分類16	
分類17	
分類18	
分類19	
分類20	
分類21	
分類22	位置繰り返し精度
分類23	〇・〇五ミリ、毎秒二・二メートル

分類テーブルによる日本語対象文固有用語の候補の分類

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.